

Harvard Journal of Law & Technology
Volume 29, Number 1 Fall 2015

**THE VALUATION OF UNPROTECTED WORKS:
A CASE STUDY OF PUBLIC DOMAIN IMAGES ON WIKIPEDIA**

Paul Heald, Kristofer Erickson[†] & Martin Kretschmer[‡]*

TABLE OF CONTENTS

I. INTRODUCTION	2
II. THEORETICAL CHALLENGES TO VALUING PUBLIC DOMAIN WORKS	5
<i>A. Differentiating the Values of Copies from Legal Rights</i>	6
<i>B. Private Value v. Social Welfare</i>	8
<i>C. Incentives and Availability</i>	10
III. VALUING PUBLIC DOMAIN IMAGES ON WIKIPEDIA: METHODOLOGY	13
IV. VALUING PUBLIC DOMAIN IMAGES ON WIKIPEDIA: FINDINGS	15
<i>A. The Public Domain and Author Pages</i>	16
<i>B. Costs Saved by Page Builders</i>	19
<i>C. Increased Traffic to Wikipedia Pages</i>	21
1. Adjusting for Popularity Based on Amazon Book Review Statistics	21
2. Matched Pairs Treatment #1 Shows 6% Traffic Increase for Authors	23
3. Matched Pairs Treatment #2 Shows 17% Traffic Increase for Authors	23
4. Matched Pairs Treatment #3 Shows 22% Traffic Increase for Composers and Lyricists	24
5. Matched Pairs Treatment #4 Shows 19% Traffic Increase for Composers and Lyricists	25
6. Controlling for Changes in Verbiage.....	25
<i>D. Extrapolating the Data to Estimate the Value of Public Domain Photographs on Wikipedia as a Whole</i>	25
1. Cost Savings on Wikipedia as a Whole.....	26
2. Increased Traffic Due to Public Domain Images on Wikipedia as a Whole	27
V. POLICY IMPLICATIONS	29
VI. CONCLUSION.....	31

I. INTRODUCTION

As a part of efforts to enforce and expand their rights, copyright owners have become adept at quantifying the collective value of the works they own.¹ They estimate for policymakers in monetary terms the value their copyrights purportedly add to the economy and the losses copyright owners suffer from infringement.² This type of evidence presents two problems. First, it erroneously conflates private gains with public gains, fallaciously linking copyright GDP to public

* Richard W. & Marie L. Corman Research Professor, University of Illinois College of Law; Fellow & Associated Researcher, CREATE, RCUK Centre for Copyright, University of Glasgow.

† Lord Kelvin Adam Smith Fellow in Social Sciences, University of Glasgow.

‡ Professor of Law, University of Glasgow; Director, CREATE, RCUK Centre for Copyright.

We would like to thank our research assistants Xiaoren Wang, Xi Zhao, Kenny Barr, Meghan Blakeley, and Andrea Wallace; and also Jeremy DeBeer, Dhammika Dharmapala, Bob Lawless, Peter Meyer, and Arden Rowell. We also appreciate the comments of Johnathan Cardy and Stephen LaPorte of the Wikimedia Foundation and of Jane Park of the Creative Commons.

1. See STEPHEN E. SIWEK, ECONOMISTS. INC., COPYRIGHT INDUSTRIES IN THE U.S. ECONOMY: THE 2013 REPORT 2 (2013), http://www.iipa.com/pdf/2013_Copyright_Industries_Full_Report.PDF [<http://perma.cc/8CHK-XB9V>] (“In 2012, the value added by the core copyright industries to U.S. GDP exceeded \$1 trillion dollars (\$1015.6 billion) for the first time, accounting for 6.48% of the U.S. economy.”); ASCAP, *ASCAP Reports Strong Revenues in 2013*, ASCAP (Feb. 12, 2014), <http://www.ascap.com/press/2014/0213-2013-financials.aspx> [<http://perma.cc/3XNL-R4QG>] (noting ASCAP distributed \$851.2 million in royalties in 2013); *Pandora Artist Payments: A View from the Artist’s Side*, SOUNDEXCHANGE BLOG (Oct. 10, 2012), <http://www.soundexchange.com/pandora-artist-payments-a-view-from-the-artists-side/> [<http://perma.cc/A3WU-N4AV>] (SoundExchange paid over \$292 million in royalties in 2011); UNESCO, CREATIVE ECONOMY REPORT 2013 SPECIAL EDITION: WIDENING LOCAL DEVELOPMENT PATHWAYS 10 (2013), <http://www.unesco.org/culture/pdf/creative-economy-report-2013.pdf> [<http://perma.cc/J5HZ-CU2L>] (“Figures . . . show that world trade of creative goods and services totaled a record US\$ 624 billion in 2011 and that it more than doubled from 2002 to 2011.”).

2. See China: Effects of Intellectual Prop. Infringement and Indigenous Innovation Policies on the U.S. Econ., Inv. No. 332-519, USITC Pub. 4226 at xiv (May 2011), <http://www.usitc.gov/publications/332/pub4226.pdf> [<http://perma.cc/C2TB-TNN9>] (claiming \$48 billion in losses due to Chinese infringement); Dina Bass, *Software Piracy Losses Jump to \$59 Billion in 2010, Report Says*, BLOOMBERG BUS. (May 12, 2011), <http://www.bloomberg.com/news/2011-05-12/software-piracy-losses-jump-to-59-billion-in-2010-report-says.html> [<http://perma.cc/4Z97-39CY>] (noting Business Software Alliance claimed \$59 billion in losses due to piracy); see also EDWARD RAPPAPORT, ECON. DIV., CONGRESSIONAL RESEARCH SERV., COPYRIGHT TERM EXTENSION: ESTIMATING THE ECONOMIC VALUES 15–17 (1998), http://digital.library.unt.edu/ark:/67531/metaacs/727/m1/1/high_res_d/98-144e_1998May11.pdf [<http://perma.cc/HLS8-ZWJM>] (estimating that failure to extend the term of U.S. copyright by twenty years would cost copyright owners \$330 million by 2017); PETERI SINERVO & TIMO E. TOIVONEN, FINNISH COPYRIGHT SOC’Y, THE CAPITAL VALUE OF COPYRIGHT ASSETS IN FINLAND 19 (2012), <http://www.cupore.fi/documents/CapitalValue.pdf> [<http://perma.cc/2276-K3AA>] (“[T]he value of copyright revenue streams in Finland were [sic] assessed to be 2 022 M € in 2008.”). See generally WORLD INTELLECTUAL PROP. ORG., GUIDE ON SURVEYING THE ECONOMIC CONTRIBUTION OF THE COPYRIGHT-BASED INDUSTRIES (2003), http://www.wipo.int/edocs/pubdocs/en/copyright/893/wipo_pub_893.pdf [<http://perma.cc/4ZWJ-39XR>].

welfare. Second, it overestimates the role of exclusive copyright in promoting the health of creative industries, as not all valuable activities in creative firms attract copyright or depend on permission from a rights holder.³ Works and ideas residing in the public domain are important sources of input to creative production. Consider, for example, a software developer who incorporates free and open source code into a new product or creates an application that is interoperable with existing information in the public domain, such as global positioning data. Further empirical research is needed to more precisely identify and quantify the non-copyright inputs in value creation. Moreover, the value of those inputs should be considered in terms of increased public welfare, not merely private income gains. This paper attempts to place a value on the contribution of public domain works to the world's largest digital encyclopedia project, Wikipedia.

Advocates of a robust public domain and restraint of copyright law extension face a challenge in estimating the value of the body of works they believe should be kept freely available.⁴ We include in our definition of the public domain: (1) works whose copyright term has expired; (2) works whose copyrights were extinguished due to the failure by their owners to observe various legal formalities; (3) works never subject to copyright because their creation pre-dated the legal recognition of copyright; (4) works expressly dedicated for free use⁵ by their authors, including U.S. government works; (5) objects without authors, such as facts; and (6) objects dedicated to the public by international agreement, such as ideas and concepts.⁶ As a matter of law, the domain of items in the above categories may be used by any member of the public without payment of a license fee.⁷

Assigning a monetary value to works in these six categories is difficult. Few would debate the value of Shakespeare's body of work, which is in the public domain, but no one has yet calculated it in economic terms. Many commentators have also made the case that copy-

3. For example, alternative valuation schemes have sought to enumerate the "fair use" economy, citing realms of economic activity, such as software coding, that overlap with traditional definitions of the creative industries. *See generally* THOMAS ROGERS & ANDREW SZAMOSSZEGI, CAPITAL TRADE, INC., FAIR USE IN THE U.S. ECONOMY: ECONOMIC CONTRIBUTION OF INDUSTRIES RELYING ON FAIR USE 7-8, 13 (2010), http://www.wired.com/images_blogs/threatlevel/2010/04/fairuseconomy.pdf [<http://perma.cc/28YU-S675>].

4. *See generally* Rufus Pollock, *The Value of the Public Domain*, INST. FOR PUB. POL'Y RESEARCH (2006), http://rufuspollock.org/papers/value_of_public_domain.ippr.pdf [<http://perma.cc/9T52-VSMP>] (arguing that the public domain has huge value without attempting to place a monetary figure on any part of it).

5. For the purposes of this paper, therefore, we include works that may be freely used under a Creative Commons license, even though in many cases the author technically retains title.

6. *See* Paul J. Heald, *The Public Domain*, in HANDBOOK ON THE ECONOMICS OF COPYRIGHT 93 (Richard Watt ed., 2014).

7. *Id.*

righted works frequently rely on some form of valuable public domain input (think of Disney's use of public domain stories and characters in *Cinderella*, *Snow White*, *Beauty and the Beast*, *Pocahontas*, etc.),⁸ but disentangling the public domain inspiration to those works and putting a monetary value on that input proves elusive.⁹ This situation creates a rhetorical imbalance, as copyright expansionists¹⁰ come to policymakers with seemingly hard figures, while public domain advocates fight back with anecdotes and intuition.¹¹

This article is one of the first attempts to quantify in monetary terms a portion of the public domain. Calculating the entire value of all public domain works is likely overly ambitious, but the value of some of its discrete corners may be partially quantified.¹² We believe that the empirical example we provide can demonstrate to policymakers more precisely how the *absence* of copyright can add economic value to a set of discrete works. In Part II, we discuss the problems economists face when trying to value both copyrighted and public domain works. So far, most attempts to place a monetary value on copyrights have focused on quantifying private value to owners (usually royalty streams) rather than net social welfare, which is the relevant touchstone for policymakers. We also explore the existing

8. See Richard A. Posner, *On Plagiarism*, ATLANTIC (Apr. 2002), <http://www.theatlantic.com/magazine/archive/2002/04/on-plagiarism/302469/> [<http://perma.cc/78BM-UGJ4>] (noting how Shakespeare, Milton, T.S. Eliot, Thomas Mann, and the makers of *West Side Story* relied on prior public domain works); Paul A. David & Jared Rubin, *Restricting Access to Books on the Internet: Some Unanticipated Effects of U.S. Copyright Legislation*, 5 REV. OF ECON. RES. ON COPYRIGHT ISSUES 23, 30 (2008) (“Disney itself has based many of its animated films on books that were in the public domain, including *Snow White and the Seven Dwarfs*, *Cinderella*, *Pinocchio*, *The Hunchback of Notre Dame*, *Alice in Wonderland*, and *The Jungle Book*, released exactly one year after Kipling’s copyrights expired.”); Chris Sprigman, *The Mouse That Ate the Public Domain: Disney, the Copyright Term Extension Act, and Eldred v. Ashcroft*, FINDLAW LEGAL NEWS (Mar. 5, 2002), http://writ.news.findlaw.com/commentary/20020305_sprigman.html [<http://perma.cc/H5FW-SMPG>]; *Recent Derivative Works Based on Frances Hodgson Burnett’s Classic The Secret Garden (1911)*, OPPOSING COPYRIGHT EXTENSION, <http://homepages.law.asu.edu/~dkarjala/OpposingCopyrightExtension/publicdomain/SecretGardenDWs.html> [<http://perma.cc/HA5P-63KW>].

9. See Pollock, *supra* note 4, at 8; GIANCARLO F. FROSIO, COMMUNIA, DELIVERABLE D1.11: FINAL REPORT 11, <http://nexa.polito.it/nexacenterfiles/D1.11COMMUNIA%20Final%20Report-nov2011.pdf> [<http://perma.cc/M268-PCJD>].

10. For a summary of international lobbying efforts, see Christopher Buccafusco & Paul J. Heald, *Do Bad Things Happen When Works Enter the Public Domain?: Empirical Tests of Copyright Term Extension*, 28 BERKELEY TECH. L. J. 1, 10–12 (2013).

11. See Pollock, *supra* note 4, at 3 (“One of the first printed texts of which we have record is a copy of the Buddhist Diamond Sutra, produced in China around 868AD. In it can be found the dedication: ‘for universal free distribution’. Clearly, the idea of public domain, that is, open access to knowledge, has been present since humanity first began to formally transmit and share ideas.”); FROSIO, *supra* note 9, at 12 (offering anecdotes about the reappearance of works by Freud, Vierne, and Capra due to their public domain status).

12. See Arden Rowell, *Partial Valuation in Cost-Benefit Analysis*, 64 ADMIN. L. REV. 723, 733–35 (2012) (arguing that partial or minimal valuations provide useful data for decision-makers, even when full valuations are impractical or impossible).

literature outlining the circumstances when public domain status increases net social welfare.

In Part III, we examine the use of public domain images (mostly photographs) on Wikipedia pages, one of the largest and most popular common-access websites, and form several hypotheses about the value the images add to the website. We ask: (1) For a given topic which spans a period of time in which images will be both in and out of copyright, are individual Wikipedia pages more likely to contain an image when public domain materials are available? (2) To what extent does the availability of public domain images lower the cost of webpage building? (3) To what extent does the addition of an image to a webpage increase traffic to that page? (4) Can the total value of both cost savings and increased traffic due to the inclusion of public domain images on Wikipedia be quantified by reference to the characteristics of a wider sample of general Wikipedia pages?

In Part IV, we develop a methodology for estimating the value added by public domain images to Wikipedia pages in terms of costs saved to the page developer and increased traffic to the page. Our study focused on two large samples of Wikipedia pages, one of authors and another of composers and lyricists. We collected, among other information, data on the birth and death dates of each subject, the date an image (if any) appeared on the subject's Wikipedia page, the legal justification for the inclusion of the image, and changes in page traffic between 2009 and 2014. We also collected pricing data from services that license digital reproductions of public domain images for online use, such as Corbis and Getty Images. Then, using a random sample of Wikipedia pages, we extrapolate our findings from the author and composer/lyricist data to the entire website. We conclude that the equivalent commercial value of public domain photos on Wikipedia is approximately \$246 to \$270 million dollars per year. In Part V, we identify some policy implications of the study with particular reference to proposed orphan works legislation.

II. THEORETICAL CHALLENGES TO VALUING PUBLIC DOMAIN WORKS

Putting a monetary value on a tangible asset, such as a copy of the ninth edition of Richard Posner's *Economic Analysis of Law* (currently \$199.50 on Amazon), is usually quite straightforward. The market price will serve as an accurate, and sometimes perfect, proxy. Calculating the value of Posner's copyright, a unique and intangible right to exclude others from copying the book, is more challenging. Calculating the value of the free availability of William Blackstone's *Commentaries on the Laws of England* (currently costless on the Project

Gutenberg website¹³ and priced at \$2.99 on Amazon), is even trickier. Before attempting to estimate the value of any item in the public domain, and before we attempt to estimate the value of public domain images on Wikipedia, we examine and account for two basic valuation challenges. In Section A, we discuss the difficulty of differentiating the value of copies from the value of legal rights, and in Section B, we explain the need to differentiate private value from social welfare. In Section C, we provide several examples of the challenges valuation poses as copyright law seeks to balance adequate creative incentives with the availability of works.

A. Differentiating the Values of Copies from Legal Rights

When valuing a copyright, economists make an essential distinction between the value of tangible works sold by the copyright owner (a book, 10 DVDs, 100 song downloads) and the right to exclude others from copying or performing the work, which is the fundamental characteristic of the copyright.¹⁴ A copyright in an artistic creation, being unique, is hard to value. When copyrights are sold, one can use the market price as an accurate proxy for value.¹⁵ In the absence of evidence from the direct sale of a copyright, economists can infer its value from royalty streams paid by licensees with varying degrees of accuracy.¹⁶ Alternatively, when a copyright is neither sold nor licensed (imagine the copyright in a self-published book), valuation can sometimes be informed by the income stream generated by sales of the tangible work because willingness to pay suggests a baseline value for a subset of buyers.

However, a valuation based on the income stream generated by sales of the tangible work protected by copyright presents two problems. First, that income stream can only be used to establish a minimum value because one typically cannot know whether the present copyright owner is the most highly valued user. The self-published book referenced above might generate far more revenue in the hands of a large traditional publisher capable of more efficiently exploiting

13. See *Commentaries on the Laws of England, Book the First* by Sir William Blackstone, PROJECT GUTENBERG (Dec. 30, 2009), <https://www.gutenberg.org/ebooks/30802> [<https://perma.cc/2AA5-KSNF>].

14. See 17 U.S.C. § 106 (2012).

15. In 1997, for example, David Bowie transferred his copyrights in twenty-five albums recorded before 1990 for a reported \$55 million. See *Bowie Bond*, INVESTOPEDIA, <http://www.investopedia.com/terms/b/bowie-bond.asp> [<http://perma.cc/36GQ-R7UQ>]. In addition, as early as 1794, the London auction house of Puttick and Simpson auctioned music copyrights and the original engraved plates essential to reproducing the works. See JAMES COOVER, *MUSIC AT AUCTION: PUTTICK AND SIMPSON (OF LONDON), 1794–1971*, 42 (1988) (sale of copyrights of the Willis Music Company).

16. See generally GORDON V. SMITH & RUSSELL L. PARR, *VALUATION OF INTELLECTUAL PROPERTY AND INTANGIBLE ASSETS* (2000).

the market. Second, calculating the portion of the profits generated by the copyright itself is difficult. Penguin, for example, publishes both copyrighted and public domain books in its selection of Penguin Classic Editions.¹⁷ A study of these “classics” showed that the public domain editions sold on average for \$9.00, while the copyrighted editions sold on average for \$14.10, suggesting an additional profit to Penguin of \$5.10 for each copyrighted volume.¹⁸ Is \$5.10 per volume the value of the Penguin copyrights? Probably not. Penguin is obligated to pay its authors a royalty for each copyrighted volume sold, which diminishes its profit margin. If the authors earn 20% per book,¹⁹ Penguin would only earn \$11.28 on sales of the copyrighted editions which is \$2.28 (or 25%) more than the public domain editions.²⁰ Nonetheless, the income represented by the 25% premium might serve as a proxy in calculating a minimum value of the copyright to the publisher.

In one important sense, the task of valuing a public domain work is easier than that of valuing a copyrighted work. As noted above, valuing a copyright requires valuing the legal right to exclude, and sales of tangible copies of the work are problematic proxies for that exclusive right. However, in the absence of any legal rights surrounding a work, value is much more directly a function of the measurable income stream a work generates. If a public domain work generates \$100,000 in profits each year for those selling it, then consumer willingness to pay can establish at least that baseline value.²¹ Unfortunately, the nature of the public domain itself complicates matters. Consider the value of the *Adventures of Sherlock Holmes*,²² one of the public domain books published by Penguin Classic Editions.²³ Pen-

17. See, e.g., PENGUIN CLASSICS, <http://www.penguin.com/static/pages/classics/hardcoverclassics/index.php> [<http://perma.cc/VW6Z-QQ95>].

18. See Paul J. Heald, *Property Rights and the Efficient Exploitation of Copyrighted Works: An Empirical Analysis of Public Domain and Copyrighted Fiction Bestsellers*, 92 MINN. L. REV. 1031, 1048–49 (2008) [hereinafter Heald, *Fiction Bestsellers*].

19. Author surveys show current author royalty rates varying from 6% to 40%, Brenda Hiatt, *Show Me the Money!: Traditional Publisher Survey*, BRENDA HIATT (Apr. 2015), <http://brendahiatt.com/show-me-the-money/traditional-publisher-survey> [<http://perma.cc/GFW3-VPXH>]. At the time the contracts with the authors of the copyrighted classics were negotiated, anecdotal evidence suggested rates around 20%.

20. See Heald, *Fiction Bestsellers*, *supra* note 18, at 1048–49.

21. Relying solely on the income stream may produce values that are both too low and too high: too low because later entrants may be more efficient at exploiting the work, so its latent value may be underestimated; and too high because some of the revenue may be driven by advantages not inherent in the work itself — the income generated by Sony’s choice of a public domain work as a new game platform will be partially driven by the value of the existing network of locked-in PlayStation users.

22. ARTHUR CONAN DOYLE, *ADVENTURES OF SHERLOCK HOLMES* (Harper & Bros. Publishers 1892) (1892).

23. See *The Adventures of Sherlock Holmes by Sir Arthur Conan Doyle*, PENGUIN BOOKS USA, <http://www.penguin.com/book/the-adventures-of-sherlock-holmes-by-arthur-conan-doyle/9780143117025> [<http://perma.cc/ZK3C-2H2M>].

guin might be coaxed into revealing the amount of profits earned by sales of its edition, but when we queried Amazon for “Sherlock Holmes,” we obtained over 5000 results. Hundreds of different publishers, not to mention movie and television show producers, have taken advantage of the public domain status of the great detective and are currently marketing thousands of versions and adaptations. As a practical matter, it is profoundly difficult to gather the information necessary to directly calculate the total amount of profits earned by a public domain book, song, or fictional character. We offer a more indirect means of measuring value in Part III below.

B. Private Value v. Social Welfare

Changes to copyright law can affect the value of both copyrighted works and works in the public domain.²⁴ An objective policymaker must determine whether society will be better off after proposed changes to the law come into effect. In doing so, the focus is on the change in the social value of copyrights, not on the value of particular copyrights to private owners. This point is made concisely by U.K. economist Rufus Pollock, who imagines a novel that initially sold for £10 in shops while it was under copyright, but that had its price reduced to £5 when it fell into the public domain and became freely available on the Internet:

Sometimes it is suggested that this results in a reduction in the value of that work for society since before the work was “worth” £10 but now is “worth” only £5 or even nothing. [This is] completely false. The value of the work has not changed at all. All that has happened is that the price has dropped. A consumer who previously valued the book at, say, £15 and who paid £10 and was left with £5 of “consumer surplus”, now pays £5 (or £0) and is left with £10 (or £15) of “surplus”.²⁵

In Pollock’s hypothetical, the copyright owner has suffered a serious loss of profits. At a minimum, it will have to lower its price by £5 in order to compete, and it will surely lose sales. Yet, society is better off because the book is still available, and at a lower price to consumers. Although difficult to quantify, the amount of consumer

24. See Matthew J. Baker & Brendan M. Cunningham, *Court Decisions and Equity Markets: Estimating the Value of Copyright Protection*, 49 J.L. & ECON. 567, 584–85 (2006) (finding that court decisions increasing the scope of copyright protection correlate with increased market capitalization of copyright owners).

25. See Pollock, *supra* note 4, at 5.

surplus resulting from the change in legal status is the value of public domain in this instance.

Unfortunately, industry estimates of copyright value provided to lawmakers typically only estimate private value.²⁶ For example, as Congress considered an additional twenty-year extension of the copyright term in 1998, the Congressional Record Service, relying on industry estimates, found that revenue to private copyright owners would decrease \$330 million through the year 2017 unless the term extension was passed.²⁷ The report found in essence that the law would increase consumer spending by \$330 million over a twenty-year period. The report contained no consideration of what consumers would gain from the legally mandated expenditure. Even copyright skeptics have made the mistake of conflating private value with social value.²⁸

An illustration might help. Suppose that a property owner owns a strip of land (distant and not visible from their residence) over which people must travel to reach a beach. The property owner charges people \$10 to cross his property to get to the beach. We could monetize his right to exclude by looking at the amount of money he gets in payments. Now suppose we want to monetize a public access easement across the property. The property owner might rightfully complain that doing so would decrease the value of his property by the market value of his right to exclude (so if 1000 people a year crossed his property, by \$10,000 each year). But it would be odd to suggest that creating a public access easement makes \$10,000 disappear from the economy. Rather, the \$10 stays in the pockets of beachgoers, who may then spend it on something else. And in addition, of course, many beachgoers who were not willing or able to pay \$10 each for access are now able to go to the beach. So the \$10,000 a year is only a very minimal measure of the amount of public benefit that accrues from the public access easement.

In 2006, two commentators did attempt to directly measure the social value of copyright in terms of consumer surplus. In order to measure the economic effect of illegal music downloading, Professors Rob and Waldfogel surveyed students at the University of Pennsylvania on their copying behaviors and how they valued particular musical works.²⁹ The authors found that illegal downloading reduced the amount of student expenditures on music by \$25 per year per student

26. See *supra* notes 1–2 and accompanying text.

27. See RAPPAPORT, *supra* note 2.

28. See Rogers & Szamoszegi, *supra* note 3, at 6 (finding that “fair use industries” such as educational institutions, software developers, and Internet search and web hosting providers generate \$4.4 trillion in revenue).

29. See generally Rafael Rob & Joel Waldfogel, *Piracy on the High C’s: Music Downloading, Sales Displacement, and Social Welfare in a Sample of College Students*, 49 J.L. & ECON. 29 (2006).

on average.³⁰ This constituted revenues lost to the record companies. They also found, however, that consumer surplus increased by \$70 per student due to the inability of the record companies to effectively price discriminate.³¹ In other words, when students were willing to pay only \$10 for an album that was priced at \$15, they illegally downloaded the album for free, generating \$10 in consumer surplus without an offsetting revenue loss to the record companies because no sale would have occurred in the absence of the download. This is a rare attempt to directly quantify consumer welfare in the copyright context.

Pollock's hypothetical and Rob and Waldfogel's study may seem to suggest that in every case the public is better off when a work becomes freely available. Such a conclusion requires two major assumptions. First, the initial term of copyright must be set long enough and the scope of protection must be robust enough to stimulate the creation of the work in the first place. If the term of protection is too short, for example, then the work may never be produced and society may be worse off. Second, the work must remain available to the public after it enters the public domain. If the lack of copyright causes the work to disappear, then the public is worse off and we should prefer prolonged copyright protection.³²

In the next section, we show a continuing need to develop techniques for quantifying the value of the public domain even when incentives and availability might be impaired by a change in legal status.

C. Incentives and Availability

The continued conflation by policymakers of private value and social welfare creates an urgent need to improve tools for quantifying the value of the public domain. As long as lobbyists assert that the size of royalty streams to private owners is a proper measure of public welfare,³³ then policymakers will need to be confronted with hard figures on the value of the public domain.

In addition, occasions may arise when copyright owners can show that a gap in protection results in the serious diminishment of incentives to create new works. For example, Rob and Waldfogel showed that on average student research participants spent \$25 less per year

30. *Id.* at 32.

31. *Id.* at 29–30.

32. *But see* Paul J. Heald, *How Copyright Keeps Works Disappeared*, 11 J. EMPIRICAL LEGAL STUD. 829, 830–31 (2014) (noting that books and music become more available to the public when they fall into the public domain). In addition, some have also argued for a third caveat, asserting that copyright protection might be necessary to prevent the “tarnishment” of the work. This is not a widely accepted argument and existing empirical work suggests tarnishment is unlikely, even when a work is associated with pornography against the wishes of its owner. *See* Chris Buccafusco & Paul J. Heald, *Testing Theories of Tarnishment* (unpublished manuscript) (on file with the authors).

33. *See* RAPPAPORT, *supra* note 2, at 5.

on music because of their illegal downloading activity.³⁴ The recording industry might be able to link that revenue loss to the public release of fewer songs and then quantify the value of the missing works.³⁵ Offsetting consumer benefit from illegal downloading quantified by Rob and Waldfogel provides relevant data to evaluations of change in copyright law designed to rebalance incentives (and may explain why a copyright-friendly Congress has failed to robustly address illegal file sharing). Since the ease of illegal downloading essentially nullifies the copyright status of a work, their study provides an instructive example of the usefulness of valuing the public domain.

Finally, copyright owners have claimed that bad things happen when works fall into the public domain,³⁶ asserting that works will disappear when they no longer have an owner.³⁷ Unavailability of works to the public would present a quantifiable social welfare problem.³⁸ In fact, even a vague estimate of unavailable works might suffice to convince policymakers to extend the term of copyright protection indefinitely because the countervailing public domain value of works that have gone missing would presumably be zero.

So far, copyright owners have been unable to demonstrate any negative effect on availability caused by the transition to public do-

34. See Rob & Waldfogel, *supra* note 29, at 32.

35. Cf. Glynn S. Lunney, Jr., *Empirical Copyright: A Case Study of File Sharing and Musical Output* 3–4 (Tulane Univ. Sch. of Law Pub. Law & Legal Theory Working Paper Series, Working Paper No. 14-2, 2014), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2372630 [<http://perma.cc/8HUR-UZWC>] (finding that the increase in file sharing and decrease in revenue led to the creation of more new music); Joel Waldfogel, *And the Bands Played On: Digital Disintermediation and the Quality of New Recorded Music* (June 25, 2012) (unpublished manuscript) http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2117372 [<http://perma.cc/RHS9-UPD9>] (empirical study showing a decline in sound recording revenues, but an increase in the production and quality of new works).

36. See Buccafusco & Heald, *supra* note 10, at 13–17.

37. See William M. Landes & Richard A. Posner, *Indefinitely Renewable Copyright*, 70 U. CHI. L. REV. 471, 475 (2003) (“[A]n absence of copyright protection for intangible works may lead to inefficiencies . . . because of impaired incentives to invest in maintaining and exploiting these works.”); Miriam Bitton, *Modernizing Copyright Law*, 20 TEX. INTELL. PROP. L.J. 65, 78 (2011) (“If [works enter] the public domain, they [become] obscure and thus no one [will] invest in them due to the problem of free riding. Items which retain enough value for future use should be given indefinite copyrights to maintain their value.”). The 1998 term extension implicitly relied on the notion that the absence of protection would result in diminished distribution and dissemination. See *Eldred v. Ashcroft*, 537 U.S. 186, 207 (2003) (concluding that Congress “rationally credited projections that longer terms would encourage copyright holders to invest in the restoration and public distribution of their works”); H.R. Rep. No. 105-452, at 4 (1998) (finding that the 1998 extension would “provide copyright owners generally with the incentive to restore older works and further disseminate them to the public”).

38. See Michael D. Smith, Rahul Telang, & Yi Zhang, *Analysis of the Potential Market for Out-Of-Print eBooks* 3 (Aug. 2012) (unpublished manuscript), <http://repository.cmu.edu/cgi/viewcontent.cgi?article=1383&context=heinzworks> [<http://perma.cc/FK63-83L2>] (finding that the failure of copyright owners to make available out-of-print books in eBook form prevents the realization of \$860 million in consumer surplus).

main status.³⁹ In fact, the opposite seems to be true. Several studies have shown that works become *more* available when they fall into the public domain:

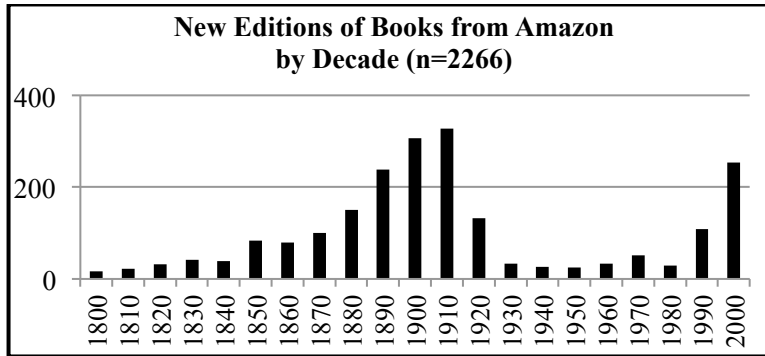


Figure 1

Figure 1 shows a large random sample of new editions of books being sold on Amazon in 2012 and organizes them by the decade of initial publication.⁴⁰ A massive increase in availability occurs for books published before 1923, the date before which all U.S. works are in the public domain.⁴¹ Another study shows that audiobooks are more likely to be made from public domain bestsellers published between 1913 and 1922 than from copyrighted bestsellers published between 1923 and 1933.⁴² Two other studies have shown that public domain music is just as likely to appear in movies as copyrighted music from the same era (1913–1932).⁴³ Public domain status does not seem to cause an availability problem.

As copyright owners continue to push for term extensions, one can see two distinct valuation issues present themselves in Figure 1. First, how great is the consumer surplus (the public domain value) embodied in the large volume of pre-1923 works that are now being offered? Second, how much does the absence of post-1923 books that have gone out-of-print diminish consumer welfare? An answer to the first question has not been attempted to our knowledge, while Professors Smith, Telang, and Zhang suggest a figure of \$860 million in

39. See Buccafusco & Heald, *supra* note 10, at 37; Heald, *supra* note 32, at 832; Heald, *Fiction Bestsellers*, *supra* note 18, at 1049–50; Paul J. Heald, *Does the Song Remain the Same? An Empirical Study of Bestselling Musical Compositions (1913–1932) and Their Use in Cinema (1968–2007)*, 60 CASE W. RES. L. REV. 1, 2 (2009) [hereinafter Heald, *Musical Compositions*].

40. See Heald, *supra* note 32, at 839.

41. *Id.* at 839–40.

42. See Buccafusco & Heald, *supra* note 10, at 22–23.

43. See Heald, *Musical Compositions*, *supra* note 39, at 10–13; Heald, *supra* note 32, at 845.

unrealized consumer surplus, as represented by books that are currently out-of-print and unexploited in eBook format by their copyright owners.⁴⁴ Finding hard numbers to better answer these questions may help policymakers more accurately predict the effect on social welfare of another round of term extensions.

III. VALUING PUBLIC DOMAIN IMAGES ON WIKIPEDIA: METHODOLOGY

Given the *Adventures of Sherlock Holmes* experiment described above,⁴⁵ we do not attempt to measure the value of public domain stories. In addition to the measurement problems caused by the existence of multiple publishers of most public domain book titles, publishers keep their revenue and sales data secret,⁴⁶ frustrating outsider attempts at valuation. Similar challenges exist when attempting to quantify the value of public domain music or film. Instead, we focus on quantifying the value of public domain images on Wikipedia, a forum which is exceedingly transparent and amenable to systematic data collection. Notably, the value that public domain images add to Wikipedia is not based on their transfer value — Wikipedia is not a market for the sale of images — so revenue streams to or from content owners, users, or the Wikipedia community need not be considered.

The valuation task, however, is hardly straightforward. Surely a Wikipedia page is more valuable if it contains an image illustrating its subject, but how much value is added? One could attempt to survey users' stated valuations, as did Rob and Waldfogel with popular music,⁴⁷ but we doubt subjects could do anything more than guess the value added by images. Most people are not familiar with market prices for online images, nor are they used to paying for access to online resources like Wikipedia. Instead, we posit that the public domain stock of images could indirectly add value in at least two ways. First, page builders save transaction costs and, potentially, licensing fees by using free images rather than negotiating with the copyright owner of an image.⁴⁸ Second, the Google search engine, which directs the vast majority of traffic on the worldwide web, prioritizes webpages with

44. See SMITH & PARR, *supra* note 16, at 3.

45. See *supra* notes 22–23 and accompanying text.

46. See Mercy Pilkington, *Anti-Transparency: Why Are Book Sales a Secret?* GOODEREADER (Feb. 19, 2014), <http://goodereader.com/blog/electronic-readers/anti-transparency-why-are-book-sales-a-secret> [<http://perma.cc/ULT2-6Q8G>] (asserting that Amazon and Barnes & Noble terms of service require that sales data be kept secret).

47. See *supra* notes 29–32 and accompanying text.

48. Wikipedia itself does not pay for permission to include images on pages, but an individual page builder can pay a copyright owner for a license to include an image.

images over pages without images;⁴⁹ therefore, Wikipedia pages with public domain images should experience more average views than pages without images. Since page visits can be valued according to the equivalent average advertising revenue generated per visitor (and a page view on Wikipedia has been estimated to have a potential value of \$.0050),⁵⁰ the value of any extra traffic driven by the image should be calculable.

Our first research question considers the scope of the effect of the reservoir of public domain images on page building: (1) Does the presence of an image on a Wikipedia page correlate with the actual historical distribution of available public domain images?

Further research questions focus on the value of an image on author, composer, and lyricist pages and the value of the set of all public domain images on Wikipedia: (2) To what extent does the availability of public domain images lower the cost of webpage building? (3) To what extent does the addition of an image to a webpage increase traffic to that page? (4) Could the total value of cost savings and increased traffic on Wikipedia be calculated by reference to the characteristics of a sample of Wikipedia pages?

Inspired by a working paper by Abhishek Nagaraj, in which he attempts to value images of baseball and basketball players on Wikipedia,⁵¹ we addressed our first research question by identifying the pages of 362 authors who had at least one bestseller on the *New York Times* bestseller's list from 1895 to 1969.⁵² These authors were born between 1829 and 1942, and their works constituted a wide mix of subjects. In the United States, all works published before 1923 are in the public domain,⁵³ so one group of authors could be represented only by a public domain image (those who died before 1923), while a second group could only be represented by a copyright-eligible photo (those born after 1923), and a third group could be represented by either a public domain or protected image (the subset of authors

49. See Abhishek Nagaraj, Does Copyright Affect Creative Reuse? Evidence from the Digitization of Baseball Digest 16–17 (Sept. 11, 2014) (unpublished manuscript), http://web.mit.edu/nagaraj/files/copyright_nagaraj.pdf [<http://perma.cc/TUC3-QAKT>].

50. See *Wikipedia*, BUS. INSIDER (Mar. 28, 2008), <http://www.businessinsider.com/companies/wikipedia> [<http://perma.cc/7AYZ-FZXG>]. A master's thesis recently estimated the value per page view on Wikipedia higher at \$.008. See Vincent Juhel, *Valorisation du benevolat sur Wikipedia* (2011) (unpublished master's thesis, HEC Paris), <http://www.amplyd.com/these/Valorisation%20du%20b%C3%A9n%C3%A9volat%20sur%20Wikip%C3%A9dia%20-%20FEB2012%20-%20Vincent%20Juhel.pdf> [<http://perma.cc/M8V9-RYC7>]. A venture capitalist has offered an even higher estimate of \$.01 per page view. See Dan Malven, *Wikipedia Should Go For-Profit, Give Profits Away*, BUSINESS INSIDER (Mar. 12, 2008), <http://www.businessinsider.com/2008/3/better-idea-for-wikipedia-goprivate-give-away-profits> [<http://perma.cc/BS5A-SB3H>].

51. See *supra* note 49.

52. See generally ALICE PAYNE HACKETT, 70 YEARS OF BEST SELLERS, 1895–1965 (1967).

53. See Buccafusco & Heald, *supra* note 10, at 6–7.

whose lives spanned the 1923 date). If author age were the only relevant factor, one might expect authors with earlier birth dates to have fewer images. After all, photographs disappear over time, so the more recent authors should have the highest percentage of images. We propose the opposite: The older the author, the more likely a public domain image will be available and the more likely an image will be used for the subject.⁵⁴

To this end we collected data on the birth and death dates of each author; the year of his or her first bestseller; the number of his or her bestsellers; the date (if any) an image of the author was added to his or her Wikipedia page; the source of the image; the legal status of the image; and the legal justification offered by the webpage builder for the use of the image.

As a measure of potential costs saved by the availability of public domain images, we searched for all the Wikipedia photographs we found in the inventory of the two largest photo licensing agencies, Corbis⁵⁵ and Getty Images,⁵⁶ and calculated the average licensing fee they charged for digital copies of photos which could be obtained freely from other sources.

In order to measure the value of potential increases in viewership due to image presence, we also counted the number of page views for each author during the months of March, April, and May 2009; the number of page views of each author during the months of March, April, and May 2014; and the number of page views during the lowest traffic month of 2009 and the lowest traffic month of 2014 for each author. In order to isolate the effect of image presence on traffic, we also collected data on changes in word count in all authors' pages from June 2009 to June 2014. We also counted the number of book reviews for each author's most-reviewed book on Amazon.

To augment our findings from author webpages, we collected a similar set of data for 792 composers and lyricists. Finally, we used the Wikipedia random page search function to generate a list of 300 random webpages in order to estimate image use and traffic on Wikipedia as a whole for the purposes of extrapolating the findings from our research on the author, composer, and lyricist webpages.

IV. VALUING PUBLIC DOMAIN IMAGES ON WIKIPEDIA: FINDINGS

We discuss below the answers to our four primary research questions. We find that the existence of a large public domain reservoir of

54. *See infra* note 59 and accompanying text.

55. CORBIS IMAGES, <http://www.corbisimages.com/> [<http://perma.cc/7TC9-T3CQ>].

56. GETTY IMAGES, <http://www.gettyimages.com/> [<http://perma.cc/2R2L-6TNK>].

photographs increases the likelihood that a webpage will contain an image and then proceed to estimate the value added by those images.

A. The Public Domain and Author Pages

The reservoir of free public domain works increases the likelihood that an author's Wikipedia page will contain an image. This is seen most clearly when considering the birth dates of the authors in our sample. All things being equal, one would assume that the pages of authors with earlier birthdates would contain relatively fewer images than those with later birthdates. A person born in 1830 should be less likely to be represented in a photograph than someone born in 1900. Photography has become cheaper and more popular over time.⁵⁷ The older a photograph, the less likely it is to survive. Our data, however, show the opposite trend in terms of inclusion of photographs on Wikipedia:

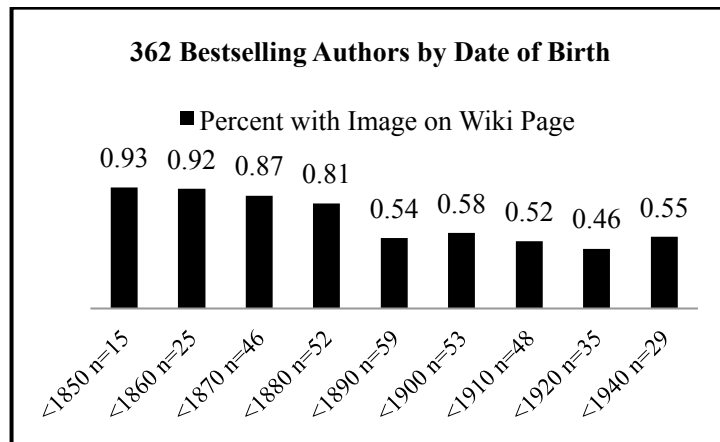


Figure 2

As the figure above clearly shows, the earlier the author's birth date, the more likely a searcher will find an image of that author on his or her Wikipedia page.

One likely reason for this trend could be reduced availability of public domain images for more recent authors due to higher transaction costs. Only half of the Wikipedia pages for 112 authors born after 1910 contain images. The image shortage almost certainly does not stem from a lack of photos of more recent authors, but probably from the higher acquisition costs associated with the copyrighted status of

57. See generally A HISTORY OF PHOTOGRAPHY: FROM 1839 TO THE PRESENT (Therese Mulligan & David Wooters eds., 2005).

the later pool of photos discussed in Section C below. We note that since our sample of authors was collected based on their book sales *at the time of publication*, the sample of older authors is not biased by any current nostalgia for those authors. In other words, there is no Darwinian effect that would lead older authors to be more image worthy.

The same trend is observed when using authors' dates of death, rather than dates of birth.

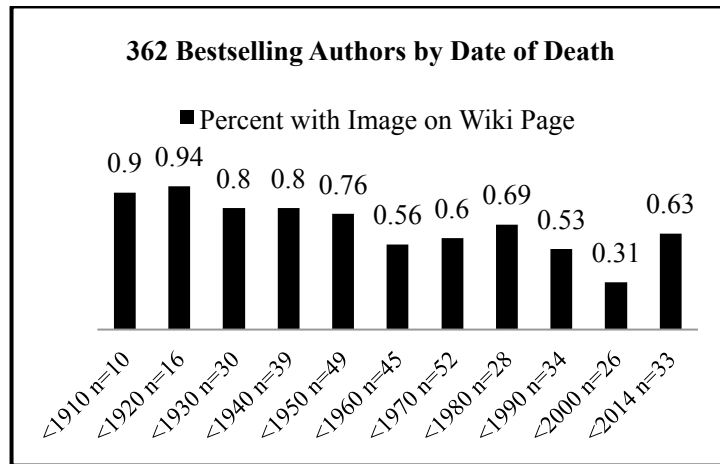


Figure 3

Page builders' reliance on the public domain is borne out by the legal status of the photos used on the author Wikipedia pages. Wikipedia requires that page builders document the source of each image and verify the legality of its use.⁵⁸ The vast majority are public domain images, although for some, fair use is claimed.⁵⁹ While some fair use is claimed, the vast majority are public domain images in the sense we use the term — works that may be freely used by Wikipedia page builders.⁶⁰

58. See, e.g., *Image of Author Stephen Crane*, WIKIMEDIA COMMONS, <http://commons.wikimedia.org/wiki/File:SCrane2.JPG> [<http://perma.cc/RWS9-C35K>] (“This media file is in the public domain in the United States. This applies to U.S. works where the copyright has expired, often because its first publication occurred prior to January 1, 1923.”).

59. Wikipedia's fair use guidance is very strict — the Wikipedia rules discourage it, probably to ensure that content remains free and open in non-U.S. jurisdictions, as well as for commercial downstream users. See *Wikipedia Image Use Policy*, WIKIPEDIA, https://en.wikipedia.org/wiki/Wikipedia:Image_use_policy [<http://perma.cc/NZ9E-VPXK>].

60. About 10% of the images we found on the author's pages were used freely by permission of an author who had granted a Creative Commons license. In such a case, the author technically retains title but grants permission to the world to freely use the image under certain circumstances, for example, with attribution given to the author. Approximate-

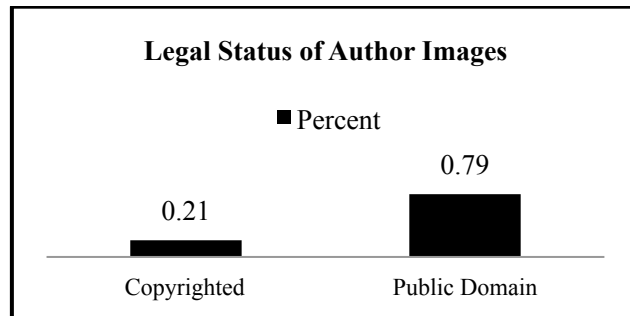


Figure 4

Wikipedia page builders typically justify their use of an image in five different ways.

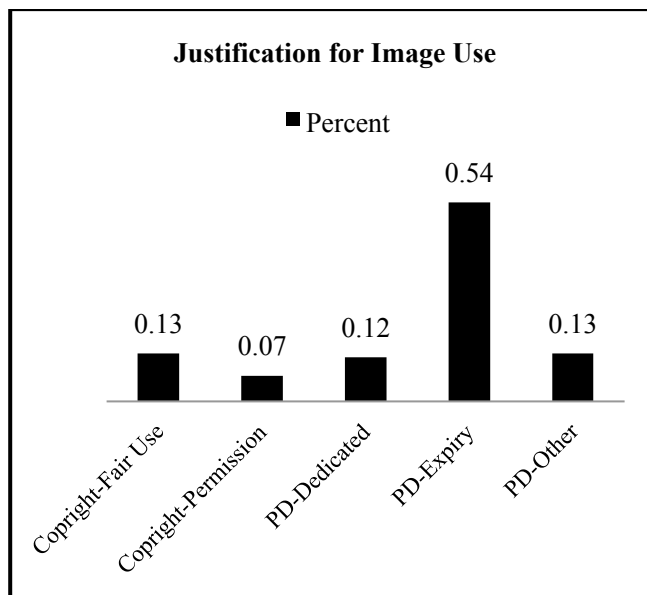


Figure 5

Most commonly, the copyright on the image has simply expired (PD-Expiry),⁶¹ while in other cases, the person taking the photograph

ly 75% of images in the Wikimedia library — an image source frequently used by page builders — are used subject to some sort of Creative Commons license. The other 25% are in the public domain due to the expiration of copyright or the failure of the copyright owner to observe some sort of formality, such as notice, registration, or renewal. *See* Email from Stephen LaPorte, Legal Counsel for Wikimedia Foundation (Jan. 24, 2015) (on file with author).

61. *See* Nagaraj, *supra* note 49.

has dedicated it for free public use, usually by referencing a form of Creative Commons license (PD-Dedicated).⁶² Some page builders take advantage of photographers who fell afoul of U.S. formalities that at one time required authors to register or renew their works or publish them with certain notice requirements (PD-Other).⁶³ Within the smaller realm of copyrighted images, the page builders typically claim fair use or obtain permission from the rights holder, otherwise they are not supposed to use a copyrighted image at all.

The existence of a large and vibrant public domain clearly increases the number of images available on author webpages. Data from random page searches supports this conclusion. Fifty percent of 300 pages collected through Wikipedia's random search function contained images.⁶⁴ Approximately 87% of the time, web builders cited the public domain as the source of an image. Approximately 8% of the time, the web builder relied on fair use of a copyrighted image, while 5% of the pages contained both copyrighted and public domain images.

B. Costs Saved by Page Builders

Webpage builders on Wikipedia save a significant amount of resources by using free public domain images. Sixty-six percent (240/362) of the author Wikipedia pages sampled contained images of the author, and 79% of those images were in the public domain. The cost savings to page builders is estimated by examining the prices for equivalent photos charged by the two largest licensors of images to webpages: Corbis Images (library of 100 million images) and Getty Images (library of 80 million images).⁶⁵ Both Corbis Images and Getty Images license images of many of the authors in this study, and sometimes they license *exactly* the same public domain image as used by Wikipedia page builders.⁶⁶

Based on price information gathered in December 2014, Corbis typically charged \$105 per year to license an image of a historically

62. See e.g., *Image of Mount Elbert*, WIKIMEDIA COMMONS, https://en.wikipedia.org/wiki/File:Mount_Elbert2.JPG [<https://perma.cc/7JZA-RSBU>] (“I, the copyright holder of this work, hereby publish it under the following licenses: This file is licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license.”).

63. See *Image of Helen MacInnes*, WIKIMEDIA COMMONS, http://commons.wikimedia.org/wiki/File:Helen_MacInnes_1969.jpg [<http://perma.cc/R8AC-87ZG>] (stating “[t]his work is in the public domain because it was published in the United States between 1923 and 1977 and without a copyright notice.”).

64. The random search function is found on the left-hand side of the Wikipedia home page, WIKIPEDIA, http://en.wikipedia.org/wiki/Main_Page [<http://perma.cc/VW42-K7PB>].

65. See *supra* notes 55–56.

66. Compare *Image of George du Maurier*, WIKIPEDIA, http://en.wikipedia.org/wiki/George_du_Maurier [<http://perma.cc/8MPR-2EQ3>], with *Image of George du Maurier*, CORBIS IMAGES, <http://www.corbisimages.com/stock-photo/rights-managed/42-17716718/portrait-of-george-du-maurier?popup=1> [<http://perma.cc/A9GQ-SFKX>].

important author for online use for a single year, while Getty regularly charged \$117 per image annually for use on a non-commercial website.⁶⁷ Curiously, in our sample, functionally identical digital versions of more than 10% (25/240) of the public domain images used on author Wikipedia pages are currently being licensed by Corbis Images or Getty Images at these rates. For 104 other public domain author images, Corbis Images or Getty Images license similar, but not identical, images of the authors. The average charge for all images was approximately \$120 per year for online usage. For the tiny slice of Wikipedia that constitutes our sample of historical authors, page builders potentially saved \$77,400 over a five-year period (129 public domain images x \$120/year x 5 years) over the cost of licensing.

Of course, this represents only a hypothetical commercial equivalent, since the \$120 average license fee is the asking price and not a direct measure of willingness to pay. Wikimedia Foundation's entire annual budget amounts to \$45.9 million,⁶⁸ and an equivalent commercial website would likely seek to avoid paying full retail price for licensed digital images by negotiating in bulk. Moreover, costs saved by page builders are not a direct measure of the value that an image creates for consumers, but they might serve as a reasonable proxy.

Using the traffic statistics function of Wikipedia,⁶⁹ we estimate that our 240 authors with images received approximately 28 million page views in 2014. What was the value to consumers of seeing images on these pages? What would they be willing to pay for an image-enhanced page? If page builders had to obtain licenses from Corbis or Getty to use these images, the total cost for the year 2014 would have been approximately \$28,000 (240 x \$120/year). This means that the per page view cost would have been about 1/10 of a penny. Would users of Wikipedia be willing to pay a penny for every 10 images they encounter on its webpages? Experimental research might best be able to elicit an answer to this question, but data collected from an advertiser's perspective may support the reasonableness of the proxy. We conclude below that the presence of an image increases page views by approximately 19%. If this is correct, then images drove 5,320,000 of our 240 author's page views in 2014. If the Business Insider estimate

67. *Compare Pricing for Image of Simone de Beauvoir*, CORBIS IMAGES, <http://www.corbisimages.com/stock-photo/rights-managed/42-58491259/simone-de-beauvoir?popup=1> [<http://perma.cc/9EX8-NDHY>] (\$105 for one-year non-commercial license), with *Pricing for Image of Simone de Beauvoir*, GETTY IMAGES, <http://www.gettyimages.com/pricecalculator/2662847> [<http://perma.cc/F96D-UQDR>] (\$117 for one-year non-commercial license).

68. Wikimedia Found., Inc., *Wikimedia Foundation Financial Statements, June 30, 2013 to June 30, 2014*, http://upload.wikimedia.org/wikipedia/foundation/b/bf/Audit_Report_-_FY_13-14_-_Final_v2.pdf [<http://perma.cc/R4WR-2755>].

69. The monthly traffic for each Wikipedia page is available at <http://stats.grok.se/> [<http://perma.cc/XLA9-LUM6>].

of a \$.0050 value for each Wikipedia page view is also correct,⁷⁰ then the advertising value of the images on our author webpages is \$26,600, almost the same as our “cost savings” estimate above.

C. Increased Traffic to Wikipedia Pages

The presence of an image is commonly believed to increase traffic to a webpage,⁷¹ but the magnitude of this effect is difficult to estimate. In general, author pages with images generate substantially more traffic than pages without images. The 240 Wikipedia author pages with images were viewed 6,764,981 times (an average of 29,000 views per page) during the months of March, April, and May of 2014, while the 122 pages without images were viewed only 385,673 times (an average of 3,260 views per page) during the same time period. Most of this difference is likely due to inequalities in the relative popularity of the authors with images. More popular authors are likely to attract more interest from web builders (and page viewers) and are more likely to have an image on their Wikipedia page. For example, the most viewed pages of authors with images were Ernest Hemingway (641,000 views during March, April, and May of 2014), F. Scott Fitzgerald (321,000 views), and Ayn Rand (301,000 views). The most popular authors without images over the same period are a less distinguished crew: Catherine Marshall (28,000 views), James Clavell (24,000 views), and Adela Johns (17,000 views).

A number of adjustments were made in order to isolate the effect of the presence of an image from the relative popularity of the authors in the study.

1. Adjusting for Popularity Based on Amazon Book Review Statistics

Author popularity was measured in terms of the number of reviews for the author’s most reviewed book on Amazon.⁷² More popular books should garner more reviews, and the market response to an

70. See *supra* note 50.

71. See Nagaraj, *supra* note 49; see also Eugene B. Visser & Melius Weideman, *An Empirical Study on Website Usability Elements and How They Affect Search Engine Optimisation*, 13 S. AFRICAN J. INFOR. MGMT. 428 (2011) (canvassing the debate over the impact of website elements on traffic); Antonio Reinoso, et al., *A Statistical Approach to the Impact of Featured Articles in Wikipedia*, <http://gsyc.es/~ajreinoso/papers/2010KEOD.pdf> [<http://perma.cc/K6UF-YHPW>] (noting featured articles — which are more likely to contain images — receive significantly more traffic on the English version of Wikipedia).

72. Unlike its sales rank data, Amazon appears to collect and organize book reviews by title rather than by edition. Willa Cather’s *My Antonia*, for example, shows 574 reviews for the paperback, hardback, and Kindle editions. See *My Antonia (Great Plains Trilogy)*, AMAZON, http://www.amazon.com/My-%C3%81ntonia-Great-Plains-Trilogy-ebook/dp/B00A73AIA2/ref=la_B00KJRO2WQ_1_1?s=books&ie=UTF8&qid=1440011267&sr=1-1 [<http://perma.cc/T84G-C4H6>]. This makes comparing review data much easier than tracking sales rank data.

author was hypothesized to be a reasonable proxy for public interest.⁷³ Authors with and without images were grouped in four categories — those with 0–9 reviews, 10–29 reviews, 30–99 reviews, and 100–199 reviews. The results show a very robust increase associated with the presence of an image.

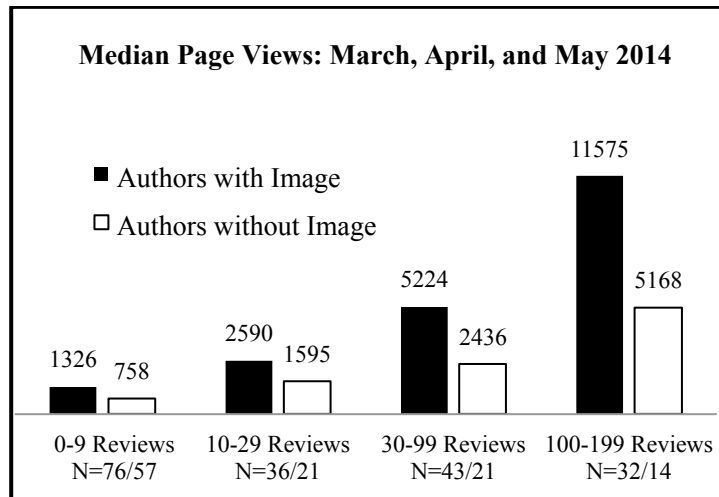


Figure 6

For example, we compared 76 authors *with* images who had fewer than 9 reviews with 57 authors *without* images who had fewer than 9 reviews. The author pages with images had on average 75% more page views in March, April, and May of 2014. This large increase associated with image use struck us as implausible and suggested that the number of Amazon reviews may be a poor proxy for author popularity. Consider Gertrude Stein (34,628 Wikipedia page views in May 2014); her most frequently reviewed book has only 49 reviews on Amazon, while the most popular work of Betty Smith (2084 Wikipedia page views in May 2014) has 1140 reviews. The Amazon adjustment for author popularity suggests that images may have *some* influence on page traffic, but we decided to employ several more sophisticated matched pairs analyses in order to better distinguish the impact of image presence from the impact of differential author popularity.

73. Using revenue data would be ideal, but those figures are proprietary. Using sales rank on Amazon as a proxy for revenue is made impossible, because many of the most popular works of the authors studied are in the public domain and therefore are represented by dozens and sometimes hundreds of different editions on Amazon stymying the estimation of overall sales. See Heald, *supra* note 32, at 840–41.

2. Matched Pairs Treatment #1 Shows 6% Traffic Increase for Authors

As a more precise measure of popularity, we identified a set of authors whose Wikipedia pages initially received an image *after* June 1, 2009, and we counted the number of views for these authors' pages for the three months immediately *prior* to June 1, 2009.⁷⁴ Each author was paired with another author of similar popularity, whose page had never contained an image. The popularity pairings were based on a comparison of the relative viewership levels during the months of March, April, and May 2009, a point in time when neither of the paired authors had an image on his or her Wikipedia page. For example, Frank H. Spearman (342 views during March, April, and May of 2009) was paired with Mary Stanberry (338 views for the same period). A total of 40 tightly matched pairs were identified, and the net increase in traffic from March, April, and May 2009, to March, April, and May 2014 was calculated.

Over the five-year period studied, the pages with images saw an increase in traffic of 32%, while the pages without images saw a net increase of only 26%. (The increase in overall traffic on Wikipedia during this time period was 22%.) The matched pairs analysis therefore showed a significantly lower net image effect (+6%) than the popularity groupings based on Amazon data presented above.

3. Matched Pairs Treatment #2 Shows 17% Traffic Increase for Authors

A second matched pairs analysis was conducted to better account for variations in web traffic caused by factors other than the addition of an image to a page. The first set of matched pairs demonstrated substantial volatility in month-to-month web traffic, indicating a variety of exogenous factors might have affected traffic levels. For example, in April 2009, Vladimir Nabokov's page was viewed 41,891 times, while the next month it was viewed 56,552 times. Schools assigning an author's book or the release of a film could easily result in short-term spikes in page visits. As a method of minimizing the impact of external factors, the lowest month of page views for the year preceding June 2009 was identified for each author. The slowest month of traffic for any author was used as a measure of the author's ambient popularity, less likely to be affected by exogenous spikes in

74. Choosing a cut-off date was driven by the need to assemble as many comparative pairs as possible. Many authors did not have pages before 2009, so choosing an earlier date would reduce the number of potential pages for analysis. Pushing the date further in the future would also reduce the number of pages eligible for analysis. For example, if we chose June 2010 instead, we would lose all pages that added an image between June 2009 and June 2010.

interest. As in the earlier matched pairs analysis, authors without images as of June 2009, were selected, and those authors with images added after June 2009 were paired with similar authors whose Wikipedia pages never contained an image. For example, Michael Gold (123 views in lowest month preceding June 2009) was paired with Harvey Wheeler (123 views in the lowest month during the same period).

The lowest page-view month in the year preceding June 2009 was compared with the lowest page-view month for the year preceding June 2014. A comparison of 42 tightly-matched pairs saw a 36% increase in traffic to the author pages containing an image, while traffic to pages without an image increased only 19% over the same five-year period. This matched pairs analysis netted a 17% increase in traffic associated with the presence of an image. As with all traffic data, any increase would be due to assumed preferences in the Google search algorithm for pages with images.

4. Matched Pairs Treatment #3 Shows 22% Traffic Increase for Composers and Lyricists

The analysis of changes in traffic to authors' webpages after making adjustments for relative author popularity netted three different increases: 100%, 17%, and 6%. Although the 17% figure generated by the second matched pairs analysis struck us as the least affected by non-image-related exogenous factors, we decided to mine a large database of well-known composers and lyricists from approximately the same era as a robustness check and to increase the number of data points.⁷⁵ We repeated both of the matched pairs techniques we used with our data set of authors.

We established 77 pairs and compared the number of page views during the period of March, April, and May 2009 before any composer or lyricist page acquired an image, with the number of page views in March, April, and May of 2014, after half of the pages acquired an image. The pairs were very tightly matched. Pages that did not acquire an image had 209,116 aggregate page views in March, April, and May of 2009, while pages that later acquired an image had 209,294 aggregate page views over the same three-month period. Between 2009 and 2014, the traffic to pages with images increased 56% while the traffic to pages without images increased only 34%, resulting in a net increase in traffic to pages with images of 22%.

Interestingly, we observed a lower level of month-to-month volatility in this data set and speculate that lower variation in month-to-month traffic may have been due to the fact that our list of composers

75. All 792 composers and lyricists were those mentioned as having hit songs from 1895–1965. *See generally* JULIUS MATTFIELD, *VARIETY MUSIC CAVALCADE* (1971).

and lyricists were less famous. For example, they averaged only half as many page views during March, April, and May of 2014 as did our list of authors.

5. Matched Pairs Treatment #4 Shows 19% Traffic Increase for Composers and Lyricists

Although the March, April, and May comparisons of page traffic on composer and lyricist webpages showed less volatility than the parallel comparison made on the author webpages, we proceeded to engage in a comparison of the lowest traffic months in 2009 and 2014 that had earlier resulted in the 17% net traffic increase figure for the authors. We were able to assemble 68 tightly matched pairs based on the lowest traffic month for each composer and lyricist in 2009 before any sample page contained an image.⁷⁶ Over the five-year period, traffic to pages with images increased 40% while the traffic to pages without images increased only 21%, resulting in a net increase of 19%.

6. Controlling for Changes in Verbiage

In order to control for the possible effect of increased verbiage on the webpages over the five-year period studied, the number of words present on all author webpages in June 2009 was compared to the number of words present on the same pages in June 2014. The change was virtually identical for the set of webpages with images and without images. Over five years, the pages with images saw an increase in word count of 66% while the pages without images saw an increase in word count of 67%. The growth of text as opposed to the addition of images does not seem to drive web traffic to the studied pages.⁷⁷

D. Extrapolating the Data to Estimate the Value of Public Domain Photographs on Wikipedia as a Whole

The analysis of a sample of 300 Wikipedia pages collected through its random search function enables us to extrapolate the author, composer, and lyricist data to Wikipedia as a whole. We offer a rough estimate of the total value of public domain photographs on Wikipedia.

⁷⁶ Due to a clerical error, the year-long period measured was June of 2009 to May of 2010, which caused us to have 9 fewer pairs than in the prior analysis which had included 77 pairs.

⁷⁷ It is possible that the number of links inserted in a page or the frequency of editing also affects web traffic to a page, but we did not measure these elements.

1. Cost Savings on Wikipedia as a Whole

Public domain photographs enable page builders to add images without incurring the cost of negotiating or paying licensing fees. A random sample we collected of 300 Wikipedia pages shows that 50% contain images, and 87% of those page builders cite the public domain as the source of the image. If the random sample is representative of Wikipedia as a whole,⁷⁸ then public domain images can be seen on 1,983,609 Wikipedia pages (4,560,201 [total Wikipedia pages as of July 18, 2014] x .50 x .87). Given that Corbis Images and Getty Images routinely charge \$105 and \$117 respectively to license a photographic image for a year on the Internet, this suggests a net savings to page builders of between \$208 million to \$232 million per year. We recognize that these figures are only a proxy for consumer surplus, but as discussed above,⁷⁹ it may be a plausible proxy. As we noted earlier, it is based on what the two largest players in the market believe they can extract from consumers, and on a per-imaged-viewed basis represents only a fraction of a cent cost per view.

Nonetheless, the estimate is rough for several reasons. In many circumstances, neither Corbis Images nor Getty Images may have an appropriate stock photo available for use on a particular page. In that case, the savings accruing to the page builder who uses a satisfactory public domain photo would best be measured in terms of the cost saved by not having to take the photo. This would vary. For example, one of the random pages is about “Netley Heath,” a rural location in England.⁸⁰ If the page builder can walk out his front door and snap a picture of the heath, then the costs saved by the existence of an easy-to-locate public domain photo would be quite small. On the other hand, if the page builder for “Netley Heath” is in the United States, the savings could be substantially larger.

Additionally, it should be noted, that active photographic opportunities avail themselves most frequently in the context of the 25% percent of Wikipedia pages about “places,” such as “Netley Heath,” or “Ely Place,”⁸¹ or the “Shudehill Interchange”⁸² (all pages from the random sample). Images for biographical pages or pages about events

78. With a sample of 300 pages, we can be confident at a 95% level that 44.3% to 55.7% of Wikipedia pages contain pictures. Assuming a total of 150 images (300 pages sampled, 150 pages with images, one image per page, which should be conservative for calculating confidence intervals), then an estimated 81.6% to 92.4% of images on Wikipedia are in the public domain.

79. See *supra* Part IV.B.

80. *Netley Heath*, WIKIPEDIA, http://en.wikipedia.org/wiki/Netley_Heath [<http://perma.cc/3AYC-HHXV>].

81. *Ely Place*, WIKIPEDIA, http://en.wikipedia.org/wiki/Ely_Place [<http://perma.cc/MXV4-J23B>].

82. *Shudehill Interchange*, WIKIPEDIA, http://en.wikipedia.org/wiki/Shudehill_Interchange [<http://perma.cc/4MN7-SZEY>].

are often impossible for a page builder to photograph. People and past events are often not available to be photographed, no matter how much the page builder is willing to spend. Among the random pages, 27% were biographical and 5% were about events in the past (for example, the “Taiyo Department Store Fire in 1973”⁸³). For the one-third of Wikipedia pages that consist of biographical or event entries, the cost savings of using a public domain photograph are best estimated by looking to licensing fees for existing photos. A final category of random pages, “things” (43% of the total), represent a mixed bag of accessibility to photographers. If one is in North Texas, it would be relatively easy to snap a photo of the “Denton County Transportation Authority.”⁸⁴ On the other hand, finding a South Asian “Banded Kingfisher”⁸⁵ willing to pose for a photograph raises greater difficulties.⁸⁶

Whether using a measure based on saved licensing fees or costs saved in locating and shooting photos, we are comfortable with estimating a cost savings in the neighborhood of \$208 to \$232 million per year based on the saved fees rationale and using that as a rough proxy for consumer surplus. Moreover, if Wikipedia were a firm, it would be reasonable to characterize these savings as producer surplus, since they arise from the lower cost of the input (images) into the product (information website).

2. Increased Traffic Due to Public Domain Images on Wikipedia as a Whole

Estimating the value added by increased traffic from public domain images is complicated by the difficulty of isolating the effect of author popularity and other exogenous factors from the effect of the addition of an image. Using the number of Amazon reviews to control for author popularity generated an increased traffic figure that seemed extremely high (over 100%), while the month-to-month volatility of the March, April, and May 2009 figures for authors also rendered its 6% finding suspect.

Our final three matched pairs analyses converge more closely. Using a lowest-month technique, we believe we were able to better control for exogenous effects on page views, and we found a traffic increase for authors of 17% and an increase for composers and lyri-

83. *Taiyo Department Store Fire*, WIKIPEDIA, http://en.wikipedia.org/wiki/1973_Taiyo_Department_Store_fire [<http://perma.cc/QDL3-FSQH>].

84. *Denton County Transportation Authority*, WIKIPEDIA, http://en.wikipedia.org/wiki/Denton_County_Transportation_Authority [<http://perma.cc/AT43-H87E>].

85. *Banded Kingfisher*, WIKIPEDIA, http://en.wikipedia.org/wiki/Banded_kingfisher [<http://perma.cc/36GJ-FQ4J>].

86. Of course, the size of licensing fees is probably affected by the work of amateur photographers who make their work available for free.

cists of 19%. Furthermore, the March, April, and May comparison for composers and lyricists obtained a similar result, a net 22% increase in traffic. We use the average of these three figures (19%) to estimate the net value of increased traffic to Wikipedia as a whole due to the widespread use of public domain images.

In order to derive a total value for increased traffic associated with the use of public domain images on Wikipedia, we multiply the total number Wikipedia pages by .5 (the percentage of pages in the random sample with images) and then by .87 (the percentage of random pages with images that rely on public domain works). We then calculate the average number of annual page views for each page with an image (18,966)⁸⁷ and credit .19 of those views to the presence of the public domain image. Finally, we multiply by the value assigned to a Wikipedia page view by Business Insider of \$.0050 per page view.⁸⁸

In total, therefore, we estimate the value of the increased traffic on Wikipedia in 2014 due to the presence of public domain images to be approximately \$35,740,074.31. $(4,560,021$ [total Wiki pages as of July 18, 2014] \times .5 [percentage of pages with images] \times .87 [percentage of pages with public domain images] \times 18,966 [average page views per year] \times .0050 [average value of a Wikipedia page view] \times .19 [percent of traffic due to public domain image] = \$35,740,074.31).

Again, we emphasize that this nearly \$36 million figure is a proxy for consumer surplus. It is more directly a measure of a premium that advertisers would be willing to pay due to traffic increases caused by the inclusion of public domain images or profits that a commercial website would be able to earn due to increased advertising revenue from visitors attracted by the presence of public domain images. Nonetheless, it captures a monetary value that Wikipedia would be able to realize were it willing to accept advertising. Given the free access and non-profit nature of Wikipedia, that surplus inures primarily to its consumers. We note that for some works not considered here, advertising revenue might well be the best proxy. Before cable television, programming was monetized exclusively through advertisements sold to those promoting consumer goods. The best way to value a television show during its run in the 1970's would be the advertising revenue it generated. Of course, the cost to consumers would have been spread nearly invisibly in terms of slightly higher prices (plus the value of their time spent watching the commercial, where that cost

87. We identified each random page with an image and counted page views for the most recent ninety-day period and multiplied by four to estimate an annual viewership for each page. The 18,966 figure is the average number of annual views per page.

88. See *supra* note 50.

exceeded the information value of the commercial), but one might fairly impute a consumer willingness to pay.

V. POLICY IMPLICATIONS

Welfare analysis in the copyright context, as widely presented, falls short of providing convincing evidence for policy decisions. Policymakers need to develop a more sophisticated understanding of non-copyright materials as both input and output, and we make a serious first attempt.

Public domain photos on Wikipedia increase net social welfare under either a cost-savings measure (\$208–\$232 million) or a traffic-increasing measure (\$38 million). One could even consider aggregating the two figures to estimate a net value for the images. Our estimate of cost savings is based on a per-web-site license price charged by Corbis Images and Getty Images. Since neither Corbis Images nor Getty Images adjust their price to account for web traffic, the savings we hypothesize do not vary with the number of page views. This means that the value of any increased traffic is mostly independent of the costs savings⁸⁹ and might be added to it in our estimate of the overall value of public domain photographs on Wikipedia, which would result in an estimate of between \$246 and \$270 million per year.⁹⁰

Since the primary purpose of this paper has been to demonstrate one possible method for estimating the value of public domain works, we will only briefly note some possible policy implications. First, we believe that the time has come for evidence-based policymaking in copyright law. Building on the Gowers Report,⁹¹ the Intellectual Property Office of the United Kingdom has already endorsed the Hargreaves Report which concludes that no further changes to U.K. intellectual property law should be made in the absence of sound empirical evidence.⁹² Econometric tools exist to help inform legislators of the

89. We suggest “mostly” because if traffic were low, then willingness to pay would drop.

90. Imagine a twenty-store bakery chain that invents a process to make a new and delicious gluten-free donut. The invention has two benefits for the bakery chain. First, it allows the chain to avoid the state gluten tax, which saves it \$1000 per year per store. Second, the new donuts are delicious, and revenues increase by \$1 million per year. The value of the invention to the bakery chain is \$20,000 + \$1 million per year. On the other hand, if the bakery chain’s savings varied with the number of extra donuts sold, then aggregating the two figures would not be appropriate.

91. See HER MAJESTY’S TREASURY, GOWERS REVIEW OF INTELLECTUAL PROPERTY, 5 (2006), <https://www.gov.uk/government/publications/gowers-review-of-intellectual-property> [http://perma.cc/YR39-QNVK] (“The Review takes an evidence-based approach to its policy analysis and has supplemented internal analysis by commissioning external experts The Review also consulted widely with a range of stakeholders in industry, academia and the public sector.”).

92. See GEORGE OSBORNE ET AL., THE GOVERNMENT RESPONSE TO THE HARGREAVES REVIEW OF INTELLECTUAL PROPERTY AND GROWTH 3 (2011), <https://www.gov.uk/>

social welfare effects of copyright regulations they consider enacting, and the time has come for a more systematized and technocratic review of the options.⁹³

More specifically, our study suggests that the most recent copyright term extension inflicted massive social harm by preventing millions of works from falling into the public domain since 1998.⁹⁴ Public domain works have a quantifiable monetary value which can be used to estimate consumer surplus. As long as the transition to public domain status does not lessen the availability of works to the public, then we find no economic justification for further retroactive extensions of copyright protection for existing works.

Finally, we conclude that our study provides a strong justification for the enactment of orphan works legislation, like that which has languished in Congress for years.⁹⁵ The proposed legislation would limit remedies to injunctive relief and a reasonable royalty when the unauthorized user of a copyrighted work cannot locate its owner after engaging in a diligent search. Similar legislation has already been passed in the United Kingdom, freeing access to an estimated 91 million works.⁹⁶

Orphan works are those that are technically protected by copyright, but their owners are unfindable by ordinary means. Around the world, the difficulty of sourcing photographs presents the most urgent case for orphan works reform.⁹⁷ Even a cursory examination of photographs in older books and magazines reveals the problem. Often no credit is given at all to a photographer or the photographer listed is long dead or cannot be located. Even when a copyright is clearly

government/publications/hargreaves-review-of-intellectual-property-and-growth-government-response [https://perma.cc/2UAB-9W5R]. (“Fundamentally, the Government agrees with not only the Review’s headline conclusion but also with its underlying critique: too many past decisions on IP have been supported by poor evidence, or indeed poorly supported by evidence.”); see also IAN HARGREAVES, DIGITAL OPPORTUNITY: A REVIEW OF INTELLECTUAL PROPERTY AND GROWTH 8 (2011), https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/32563/ipreview-finalreport.pdf [https://perma.cc/UU8W-R62G] (“Government should ensure that development of the IP System is driven as far as possible by objective evidence. Policy should balance measurable economic objectives against social goals and potential benefits for rights holders against impacts on consumers and other interests.”).

93. See Shawn Bentley Orphan Works Act of 2008, S. 2913, 110th Cong. § 514 (2008) (proposing injunctive relief and reasonable compensation only when a copyrighted work is exploited after a good faith search has failed to locate its owner).

94. See *Statement of Marybeth Peters the Register of Copyrights Before the Subcomm. on Cts., the Internet, & Intell. Prop. Comm. on the Judiciary*, 110th Cong. (2008), <http://www.copyright.gov/docs/regstat031308.html> [http://perma.cc/42EL-9FQR] [hereinafter *Statement of Marybeth Peters*] (noting the special problems posed by photographs).

95. See S. 2913, *supra* note 93.

96. See Press Release, Dep’t for Bus., Innovation & Skills, Intell. Prop. Off., Baroness Neville-Rolfe DBE CMG., UK Opens Access to 91 Million Orphan Works (Oct. 29, 2014), <https://www.gov.uk/government/news/uk-opens-access-to-91-million-orphan-works> [https://perma.cc/9MVR-QJJ6].

97. See *Statement of Marybeth Peters*, *supra* note 94.

claimed, it is often impossible for a potential user/licensee to determine whether it was properly renewed and is therefore still protected. Many photographs, perhaps even the vast majority, were never registered or renewed at all.

Imagine a photograph of President Franklin Roosevelt in a 1935 newspaper. In order to be protected by copyright the photo must have been registered and then renewed in 1963, but the Copyright Office website does not allow users to limit their searches to only photographs. And finding a file entry entitled “Franklin Roosevelt” on the website does not ensure that the newspaper photograph sought to be used is the one referenced in the registration. Even a visit to the copyright office in Washington, D.C. (which will be necessary because renewal records are not online) will not enable the potential user to leaf successfully through all registered photographs of Franklin Roosevelt because changes to the law that formerly required deposit means that not all registered copies of photographs are available in the catalog.

Given the demonstrated value of public domain photographs on Wikipedia, millions of more valuable photographs might join the party if page builders were not intimidated by the danger of incurring infringement damages in the numerous cases where the copyright owner cannot be tracked down or the renewal status of the work cannot be confirmed. Some version of orphan works legislation would clearly increase consumer surplus, and thus improve social welfare.

VI. CONCLUSION

In the absence of established market prices, valuation is always the domain of estimation and proxies. This is especially true of intellectual property in copyrights and patents, where works are original or novel by definition. Nevertheless, the exercise of quantifying the value of legal rights, *and the value of the absence of legal rights*, illuminates issues for policymakers even when precise numbers cannot be put on consumer surplus and overall social welfare. Our study demonstrates that the value of the public domain can be estimated at least as precisely as the commercial value of copyrights. Even though our estimates make use of several proxies, implications for both copyright term extension and orphan works legislation are substantial. The time has come for the Copyright Office and the U.S. Congress to endorse an evidence-based regime for the federal management of creative works.